Machine Learning for Tomographic Imaging

Ge WANG, PhD; wangg6@rpi.edu
Biomedical Imaging Center
AI-based X-ray Imaging System (AXIS) Lab
CBIS/BME, Rensselaer Polytechnic Institute

University of Minnesota Institute of Mathematics, October 16, 2019
Machine Learning for Tomographic Imaging

Göran Wang, PhD; wangg6@rpi.edu
Biomedical Imaging Center
AI-based X-ray Imaging System (AXIS) Lab
CBIS/BME, Rensselaer Polytechnic Institute

October 16, 2019
Machine Learning for Tomographic Imaging

Gø WGAN, PhD; wangg6@rpi.edu
Biomedical Imaging Center
AI-based X-ray Imaging System (AXIS) Lab
CBIS/BME, Rensselaer Polytechnic Institute

October 16, 2019
ACKNOWLEDGMENT

The author is grateful for inspiring discussions with Drs. James Brink (Harvard MGH), Wenxiang Cong (RPI), Juergen Hahn (RPI), Michael Insana (UIUC), Nadeem Ishaque (GE GRC), Mannudeep K. Kalra (Harvard MGH), Xuanqin Mou (Xi’an Jiaotong Univ., China), Jiantao Pu (Univ. of Pittsburg), Dinggang Shen (Univ. of North Carolina), Michael Vannier (Univ. of Chicago), Yan Xi (RPI), George Xu (RPI), Pingkun Yan (Philips), Hengyong Yu (Univ. of Massachusetts Lowell), Junping Zhang (Fudan Univ., China), Yi Zhang (Sichuan University), and other colleagues who made valuable comments. He also acknowledges the advice from an anonymous reviewer who gave advice on the theoretical limitations of compressed sensing. Figures 7, 8 and 9 were produced by Mr. Qingsong Yang (RPI).
A Perspective on Deep Imaging

GE WANG (Fellow, IEEE)
Department of Biomedical Engineering, Biomedical Imaging Center, Center for Biotechnology and Interdisciplinary Studies, Rensselaer Polytechnic Institute, Troy, NY 12180, USA
Corresponding author: G. Wang (ge-wang@ieee.org)

ABSTRACT The combination of tomographic imaging and deep learning, or machine learning in general, promises to empower not only image analysis but also image reconstruction. The latter aspect is considered in this perspective article with an emphasis on medical imaging to develop a new generation of image reconstruction theories and techniques. This direction might lead to intelligent utilization of domain knowledge from big data, innovative approaches for image reconstruction, and superior performance in clinical and preclinical applications. To realize the full impact of machine learning for tomographic imaging, major theoretical, technical and translational efforts are immediately needed.
Image Reconstruction Is a New Frontier of Machine Learning

Ge Wang, Fellow, IEEE, Jong Chu Ye, Senior Member, IEEE, Klaus Mueller, Senior Member, IEEE, and Jeffrey A. Fessler, Fellow, IEEE

I. INTRODUCTION

Over the past several years, machine learning, or more generally artificial intelligence, has generated overwhelming research interest and attracted unprecedented public attention. As tomographic imaging researchers, we share the excitement from our imaging perspective [item 1] in the Appendix, and organized this special issue dedicated to the theme of “Machine learning for image reconstruction.” This special issue is a sister issue of the special issue published in May 2016 of this journal with the theme “Deep learning in medical imaging” [item 2] in the Appendix. While the previous special issue targeted medical image processing/analysis, this special issue focuses on data-driven tomographic reconstruction. These two special issues are highly complementary, since image reconstruction and image analysis are two of the main pillars for medical imaging. Together we cover the whole workflow of medical imaging: from tomographic raw data/features to reconstructed images and then extracted diagnostic features/readings.
Acknowledgment

We are supported by NIH, General Electric, Hologic, IBM, & NVIDIA. We are seeking collaborative opportunities.
Outline

- Theoretical Exploration
- Translational Efforts
- Teaching Innovation
Various Types of Cells

- Stem Cells
- Bone Cells
- Blood Cells
- Muscle Cells
- Fat Cells
- Skin Cells
- Nerve Cells
- Endothelial Cells
- Sex Cells
- Pancreatic Cells
- Cancer Cells

Types of Neurons:
- Bipolar (Interneuron)
- Unipolar (Sensory Neuron)
- Multipolar (Motoneuron)
- Pyramidal Cell
Quadratic Neurons as Fuzzy Gates & Factors
Fuzzy Logic Interpretation of Artificial Neural Networks

Fenglei Fan, Student Member, IEEE, Ge Wang, Fellow, IEEE

Abstract — Over past several years, deep learning has achieved huge successes in various applications. However, such a data-driven approach is often criticized for lack of interpretability. Recently, we proposed artificial quadratic neural networks consisting of second-order neurons in potentially many layers. In each second-order neuron, a quadratic function is used in the place of the inner product in a traditional neuron, and then undergoes a nonlinear activation. With a single second-order neuron, any fuzzy logic operation, such as XOR, can be implemented. In this sense, any deep network constructed with quadratic neurons can be interpreted as a deep fuzzy logic system. Since traditional neural networks and second-order counterparts can represent each other and fuzzy logic operations are naturally implemented in second-order neural networks, it is plausible to explain how a deep neural network works with a second-order network as the system model. In this paper, we generalize and categorize fuzzy logic operations implementable with individual second-order neurons, and then perform statistical/information theoretic analyses of exemplary quadratic neural networks.

were identified [8]. However, these results do not reveal the inner working of a network, such as what and how features are extracted and propagated between layers. Gu et al. 2017 [9] offered an elegant explanation of the adversarial mechanism of GAN from the viewpoint of optimal mass transportation. Dong et al. 2017 [10] established a correspondence between deep networks and numerical ordinary differential equations to guide the structural design of a network with skip connections.

Instead of handling with existing models directly, researchers also tried to find the models that are more interpretable. For example, Wu et al. [11] utilized tree regularization to optimize a deep model with more interpretability. Fan [12] proposed a generalized hamming network based on the fact that neurons calculate generalized hamming distance when a bias is adopted. Albeit novel and interesting models developed in these pilot studies, these arts do not reveal the key mechanism based on which the existing models are so successful.
Fuzzy Logic System

https://microcontrollerslab.com/fuzzy-logic-system-working-example/
Deep Fuzzy Logic System

Fig. 3. Second-order network for recognition of Arabic digits from the MNIST dataset. The neurons in the “convolutional” layers are colored and inflated to demonstrate the types and frequencies of quadratic operations.
From Data-driven to Rule-based
Universal Approximation with Quadratic Deep Networks

Fenglei Fan\textsuperscript{a}, Jinfur Xiong\textsuperscript{b} and Ge Wang\textsuperscript{a,\textdagger}

\textsuperscript{a}Department of Biomedical Engineering, Rensselaer Polytechnic Institute, Troy, NY, 12180 USA
\textsuperscript{b}IBM Thomas J. Watson Research Center, Yorktown Heights, NY, 10598, USA

\textbf{ARTICLE INFO}

Keywords:
deeplearning
quadraticnetworks
approximationtheory

\textbf{ABSTRACT}

Recently, deep learning has achieved huge successes in many important applications. In our previous studies, we proposed quadratic/second-order neurons and deep quadratic neural networks. In a quadratic neuron, the inner product of a vector of data and the corresponding weights in a conventional neuron is replaced with a quadratic function. The resultant quadratic neuron enjoys an enhanced expressive capability over the conventional neuron. However, how quadratic neurons improve the expressing capability of a deep quadratic network has not been studied up to now, preferably in relation to that of a conventional neural network. Regarding this, we ask four basic questions in this paper: (1) for the one-hidden-layer network structure, is there any function that a quadratic network can approximate much more efficiently than a conventional network? (2) for the same multi-layer network structure, is there any function that can be expressed by a quadratic network but cannot be expressed with conventional neurons in the same structure? (3) Does a quadratic network give a new insight into universal approximation? (4) To approximate the same class of functions with the same error bound, is a quantized quadratic network able to enjoy a lower number of weights than a quantized conventional network? Our main contributions are the four interconnected theorems shedding light upon these four questions and demonstrating the merits of a quadratic network in terms of expressive efficiency, unique capability, compact architecture and computational capacity respectively.
The Power of Depth for Feedforward Neural Networks

Ronen Eldan
Weizmann Institute of Science
ronen.eldan@weizmann.ac.il

Ohad Shamir
Weizmann Institute of Science
ohad.shamir@weizmann.ac.il

Abstract

We show that there is a simple (approximately radial) function on $\mathbb{R}^d$, expressible by a small 3-layer feedforward neural networks, which cannot be approximated by any 2-layer network, to more than a certain constant accuracy, unless its width is exponential in the dimension. The result holds for virtually all known activation functions, including rectified linear units, sigmoids and thresholds, and formally demonstrates that depth – even if increased by 1 – can be exponentially more valuable than width for standard feedforward neural networks. Moreover, compared to related results in the context of Boolean functions, our result requires fewer assumptions, and the proof techniques and construction are very different.

Theorem 1. Suppose the activation function $\sigma(\cdot)$ satisfies assumption $[\text{I}]$ with constant $c_\sigma$, as well as assumption $[\text{II}]$. Then there exist universal constants $c, C > 0$ such that the following holds: For every dimension $d > C$, there is a probability measure $\mu$ on $\mathbb{R}^d$ and a function $g : \mathbb{R}^d \rightarrow \mathbb{R}$ with the following properties:

1. $g$ is bounded in $[-2, +2]$, supported on $\{x : \|x\| \leq C\sqrt{d}\}$, and expressible by a 3-layer network of width $C_{c_\sigma}d^{19/4}$.

2. Every function $f$, expressed by a 2-layer network of width at most $cC_{c_\sigma}d$, satisfies

$$\mathbb{E}_{x \sim \mu} (f(x) - g(x))^2 \geq c.$$
The Expressive Power of Neural Networks: A View from the Width

Zhou Lu, Hongming Pu, Feicheng Wang, Zhiqiang Hu, Liwei Wang

(Submitted on 8 Sep 2017 (v1), last revised 1 Nov 2017 (this version, v3))

Theorem 4. Let $n$ be the input dimension. For any integer $k \geq n + 4$, there exists $F_{\mathcal{A}} : \mathbb{R}^n \to \mathbb{R}$ represented by a ReLU neural network $\mathcal{A}$ with width $d_m = 2k^2$ and depth $h = 3$, such that for any constant $b > 0$, there exists $\epsilon > 0$ and for any function $F_{\mathcal{B}} : \mathbb{R}^n \to \mathbb{R}$ represented by ReLU neural network $\mathcal{B}$ whose parameters are bounded in $[-b, b]$ with width $d_m \leq k^{3/2}$ and depth $h \leq k + 2$, the following inequality holds:

$$\int_{\mathbb{R}^n} (F_{\mathcal{A}} - F_{\mathcal{B}})^2 \, dx \geq \epsilon.$$  

(6)
Width: n+4 & Depth: Deep

Figure 1: One block to simulate the indicator function on $[a_1, b_1] \times [a_2, b_2] \times \cdots \times [a_n, b_n]$. For $k$ from 1 to $n$, we "chop" two sides in the $k$th dimension, and for every $k$ the "chopping" process is completed within a 4-layer sub-network as we show in Figure 1. It is stored in the $(n+3)$th node as $L_1$ in the last layer of a $d$-cube. Then we use a single layer to record it in the $(n+1)$th or the $(n+2)$th node, and reset the last two nodes to zero. Now the network is ready to simulate another $(n+1)$-dimensional cube.
Deep Depth: Univariate Polynomial of Order N
Log Operation: Loudness & Brightness in dB

Equal-loudness contours (red) (from ISO 226:2003 revision)
Original ISO standard shown (blue) for 40-phon
Wide Width: Univariate Polynomial of Order N

Quadratic Neurons as Factors (with Real coefficients, According to the Algebraic Fundamental Theorem)

\[ n_1(x) = \text{Relu}(-5x - 7.7) \]
\[ n_2(x) = \text{Relu}(-1.2x - 1.3) \]
\[ n_3(x) = \text{Relu}(1.2x + 1) \]
\[ n_4(x) = \text{Relu}(1.2x - .2) \]
\[ n_5(x) = \text{Relu}(2x - 1.1) \]
\[ n_6(x) = \text{Relu}(5x - 5) \]

\[ Z(x) = -n_1(x) - n_2(x) - n_3(x) + n_4(x) + n_5(x) + n_6(x) \]

So visually at least, it looks like it IS possible to model non-trivial functions with a single hidden layer and a handful of neurons. Pretty cool.
• Duality in Depth & Width of the Architecture
• Depth & Width Convertible
• Diverse Networks for the Same Task
• Same Network Performs Different Tasks
Particle/Factor Mathematics

\[ f(x) = f(x_1, \ldots, x_n) = \sum_{q=0}^{2n} \Phi_q \left( \sum_{p=1}^{n} \phi_{q,p}(x_p) \right) \]

Also, aided by the concept of partially separable functions, the complexity of the quadratic network can be further reduced, such as in the case of computing an \( L^{th} \) separable function. By the \( L^{th} \) separable function, we mean that
\[ f(x_1, \ldots, x_n) \] is \( L^{th} \) separable defined as follows:

\[ f(x_1, \ldots, x_n) = \sum_{l=1}^{L} \prod_{i} \phi_{li}(x_i). \]

In practice, almost all continuous functions can be represented as \( L^{th} \) separable functions, which are of low ranks at the same time.

Kolmogorov
Outline

- Theoretical Exploration
- Translational Efforts
- Teaching Innovation
Low-dose CT via convolutional neural network

Hu Chen,1,2 Yi Zhang,1,* Weihua Zhang,1 Peixi Liao,3 Ke Li,1,2 Jiliu Zhou,1 and Ge Wang4

* Author information • Article notes • Copyright and License information Disclaimer
WGAN Network for Low-dose CT

Low-quality CT Scan (1/4 Radiation Dose)  High-quality CT Scan (Standard Radiation Dose)  Machine Learning Turns Low-quality CT Image into High-quality Counterpart

GE Healthcare for Low-dose CT (RSNA’18)
Commercial Iterative Recon (IR) Algorithms in This Study

Our MAP Network-based Deep Learning (DL) with Optimized Depth
Sparse-data CT De-artifacts: “LEARN”

Under-sampled Data

Reconstruction

Iteration-inspired Layer

LEARN: Learned Experts’ Assessment-based Reconstruction Network for Sparse-data CT. IEEE Trans. Medical Imaging, June 2018
AUTOMAP with $O(N^4)$ Complexity

$O(N^3 \times N_v)$ Complexity

$N$: # of Data or Pixels per Row

$N_v$: # of views

3 x 3 k-space Data

Reshape To a Vector

FC Layers to All Pixels

3 x 3 Image
Intelligent CT Network (iCT-Net)

PROCEEDINGS OF SPIE

Yinsheng Li, Ke Li, Chengzhu Zhang, Juan Montoya, Guang-Hong Chen, "Image reconstruction from fully-truncated and sparsely-sampled line integrals using iCT-Net," Proc. SPIE 10948, Medical Imaging 2019: Physics of Medical Imaging, 109480S (7 March 2019); doi: 10.1117/12.2513088

Event: SPIE Medical Imaging, 2019, San Diego, California, United States
iCTNet with $O(N^2 \times N_c)$ Complexity

Huidong Xie, Hongming Shan, Wenxiang Cong, Xiaohua Zhang, Shaohua Liu, Ruola Ning, Ge Wang, "Dual network architecture for few-view CT - trained on ImageNet data and transferred for medical imaging," Proc. SPIE 11113, Developments in X-Ray Tomography XII, 111130V (10 September 2019); doi: 10.1117/12.2531198

Event: SPIE Optical Engineering + Applications, 2019, San Diego, California, United States
Few-view CT

Few-view Sinogram → Intermediate FBP Image → FBPFConvNet → Reconstructed Image

Figure 1. Workflow of the proposed method. Images are example outputs from a 49-view sinogram. The display window
DNA: Generator 1

Filtration → Backprojection → Refinement
Backprojection with $O(N)$ Complexity

Smart Backpropagation with Learned Weights Restricted to a Single X-ray Path
Backprojection with $O(N^2)$ Complexity

Smart Backpropagation with Learned Weights Restricted to a Single X-ray Path

Learned Filtered Projection
Comparative Study


Koning Breast CT via Deep Recon

Al-based Breast Slices at 1/3 X-ray Dose

Current Breast Slices at 1/3 X-ray Dose

Current Breast Slices at Full X-ray Dose
Outline

- Theoretical Exploration
- Translational Efforts
- Teaching Innovation
Medical Imaging Course at RPI

Ge Wang
Director of Biomedical Imaging Center (BIC) & AI-based X-ray Imaging System (AXIS) Lab, Rensselaer Polytechnic Institute
## Schedule for Fall 2019

<table>
<thead>
<tr>
<th>Tue</th>
<th>Topic</th>
<th>Fri</th>
<th>Topic</th>
</tr>
</thead>
<tbody>
<tr>
<td>08/30</td>
<td>Introduction</td>
<td>09/03</td>
<td>Monday Schedule</td>
</tr>
<tr>
<td>09/06</td>
<td>Hands-on 1: Colab &amp; MNIST</td>
<td>09/10</td>
<td>Network &amp; Backpropagation</td>
</tr>
<tr>
<td>09/13</td>
<td>Hands-on 2: MNIST Finish</td>
<td>09/17</td>
<td>Architectures &amp; Interpretability</td>
</tr>
<tr>
<td>09/20</td>
<td>Image Quality: Resolution, Noise, &amp; SSIM</td>
<td>09/24</td>
<td>CT Physics (Recorded)</td>
</tr>
<tr>
<td>09/27</td>
<td>CT Reconstruction</td>
<td>10/01</td>
<td>Deep CT Reconstruction</td>
</tr>
<tr>
<td>10/04</td>
<td>Hands-on 3: CT Networks</td>
<td>10/08</td>
<td>More CT Stuff</td>
</tr>
<tr>
<td>10/11</td>
<td>MRI Physics</td>
<td>10/15</td>
<td>Spin-Echo</td>
</tr>
<tr>
<td>10/18</td>
<td>K-Space &amp; More</td>
<td>10/22</td>
<td>Review, Quiz, Q &amp; A</td>
</tr>
<tr>
<td>10/25</td>
<td>Deep MRI Reconstruction</td>
<td>10/29</td>
<td>Hands-on 4: MRI Networks</td>
</tr>
<tr>
<td>11/01</td>
<td>Nuclear Physics</td>
<td>11/05</td>
<td>Nuclear Imaging</td>
</tr>
<tr>
<td>11/08</td>
<td>US Imaging</td>
<td>11/12</td>
<td>Optical Imaging</td>
</tr>
<tr>
<td>11/15</td>
<td>Multimodality Imaging</td>
<td>11/19</td>
<td>Image Quality: Task-specific</td>
</tr>
<tr>
<td>11/22</td>
<td>Q &amp; A</td>
<td>11/26</td>
<td>Hands-on 5: More</td>
</tr>
<tr>
<td>11/29</td>
<td>Nov. 27-29 Thanksgiving</td>
<td>12/03</td>
<td>Q &amp; A</td>
</tr>
<tr>
<td>12/06</td>
<td>Project Presentation</td>
<td>12/10</td>
<td>Academic/Industrial Outlook</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Dec. 11 Last Day</td>
</tr>
</tbody>
</table>
First Book on Deep Recon (eBook & Hardcopy)
Theme & Coauthors

Image Analysis, One of Successful Applications of Artificial Intelligence & Machine Learning

Tomographic Data Acquisition

Data As Tomographic Features

Reconstructed Image

Theme of This Book:
Tomographic Image Reconstruction

Theme of This Book:
Tomographic Image Reconstruction

Coauthors of the Book:
Yi Zhang, Xiaojing Ye, Xuanqin Mou
Outline of Our Book

Part I
Basic
- Chapter 1: Background Knowledge
  - Imaging Principles, Prior Info, Vision System, Sparse Coding

Part II
CT
- Chapter 4: X-ray Computed Tomography
  - Data Acquisition, Analytic & Iterative Recon, Scanner
- Chapter 5: Deep CT Reconstruction
  - Image & Data Domain, Hybrid, Unrolled/Inserted, Direct Networks

Part III
MRI
- Chapter 6: Magnetic Resonance Imaging
  - MRI Physics, CS-based Recon, Parallel MRI
- Chapter 7: Deep MRI Reconstruction

Part IV
Others
- Chapter 8: Modalities & Integration
  - Nuclear Tomography, Ultrasound & Optical Imaging, Multimodality Imaging
- Chapter 9: Image Quality Assessment
  - General, Specific, Task-based Metrics, Network-based Observers
- Chapter 10: Quantum Computing
  - Wave-Particle Duality, Quantum Gates, Quantum Algorithms

Part V
Appendices
- Appendix A: Mathematical Basics
  - Optimization, Inference, Information Theory
- Appendix B: Hands-on Experience
  - Basic Networks, Deep CT & MRI Networks
Recognizing Your Digits

- Read an image file
- Normalize it
- Resize to a desired size
- Pass into the network
GAN: One of Greatest Ideas
Optimal Discriminator

If a new datum, if discriminator function is greater than 0.5, the datum should belong to Class 1 (blue); otherwise, it comes more likely from Class 2 (green)
Optimal Generator

Illustration of the inverse transform method. In blue: the uniform distribution over [0,1]. In orange: the standard gaussian distribution. In grey: the mapping from the uniform to the gaussian distribution (inverse CDF).

Adversarial Learning

Data/Frequency

Synthetic/Source/Fake Distribution

Data/Target/Real Distribution

Age manifold

Image space
Adversarial Learning: Initial Interplay

Data

Likelihood

Frequency

0.5 0.5

Improve
Generator

Improve
Discriminator
Adversarial Learning: Further Improvement

Data Likelihood

Frequency

0.5 1.0

Improve Generator

Improve Discriminator

Data

1.0

0.5
Simultaneous CT-MRI

Joint CT-MRI Reconstruction

Simultaneous CT-MRI Scan

Few-view CT Scan

Fast MRI Scan


Deformation Map

MA: CT

Patches

Update

Deformed Atlas

Current CT Image

MA: MRI

Patches

Update

Deformed Atlas

Current MRI Image

Fused CT-MRI Image

Ensemble Learning Improves MRI Resolution
Dynamic Cardiac MRI

Sunnybrook Cardiac Data
Normal (35 slices, 20 phases)
Cine steady state free precession (SSFP) MR short axis (SAX) images
- Slice thickness=8mm
- Gap=8mm
- FOV=320mm by 320mm
- Matrix=256 by 256
- 1.5mm x 1.5 mm resolution
Excitement = MRI & CT Coupled ($E = MC^2$)